

关于云计算可用性的定性与定量研究

(A Qualitative and Quantitative Study on Availability of Cloud Computing)

(第五部分)

陈怀临, 弯曲评论创办人
北极光创投投资顾问, 云基地中云网技术顾问

Email: huailin@gmail.com

4 案例研究--亚马逊AWS



















4.2 Amazon AWS服务宕机调查 (2006-2009)

Amazon AWS自2006年3月14开放S3文件存储服务和2006年8月25日的EC2服务, 2008年8月的EBS服务以来, 经历过许多服务崩溃下线. 其中包括EC2, S3和EBS等. 其影响面涉及到租用其服务的许多重要的互联网公司.

AWS在2008年2月16日, AWS的S3发生严重的服务宕机并导致许多AWS的用户的服务中断. Amazon的AWS团队进行了深刻的反思, 并在4月8日, 开始提供AWS Service Health Dashboard, 每天跟踪发布各种服务的可靠性.

Apr 14, 2008

[Report an Issue](#)

Current Status	Details	RSS
 Amazon Elastic Compute Cloud (API)	Service is operating normally.	
 Amazon Elastic Compute Cloud (Instances)	Service is operating normally.	
 Amazon Flexible Payments Service	Service is operating normally.	
 Amazon Mechanical Turk (Requester)	Service is operating normally.	
 Amazon Mechanical Turk (Worker)	Service is operating normally.	
 Amazon SimpleDB	Service is operating normally.	
 Amazon Simple Storage Service (EU)	Service is operating normally.	
 Amazon Simple Storage Service (US)	Service is operating normally.	
 Amazon Simple Queue Service	Service is operating normally.	

本节试图对AWS上线的重大宕机事件做一个整理列表,并做相应的讨论.

1 Apri 1, 2006

Amazon在开放其S3存储服务不到一个月,在2006年4月1日,S3发生宕机事件.

事故原因:S3


事故恢复:6个小时

事故解释: AWS团队做S3存储的负载均衡的管理调配.结果导致内部网络负载崩溃,从而使得S3子系统服务宕机.


相关URL: <https://forums.aws.amazon.com/thread.jspa?threadID=10185>

Re: Http/1.1 Service Unavailable - Outage?

 Reply

Posted by:  Dave Barth REAL NAME™

Posted on: Apr 4, 2006 5:38 PM

 in response to: Dave Barth

Amazon S3 Customers,

Below is an update on the issues causing the Amazon S3 outage last Saturday evening, which we have isolated and resolved.

We were taking the low-load Saturday as an opportunity to perform some maintenance on the storage system, specifically on some very large (>100 million objects) buckets in order to obtain better load-balancing characteristics. Normally this procedure is entirely transparent to users and bucket owners. In this case, the re-balancing caused an internal transit link to become flooded, this cascaded into other network problems, and the system was made unavailable.

We are taking several steps to ensure that we don't run into this situation again. We are modifying our maintenance procedures, and are adding further monitoring to prevent the transit link from getting full. In addition we are modifying the way that our system makes use of the network to prevent the cascading effect we saw on Saturday.

Providing world-class reliability is our top priority for Amazon S3. We appreciate your patience, and hope to surpass your expectations going forward.

-dave

2. Sept 29 . 2007

Amazon的EC2发生宕机, 有些客户丢失了数据. EC2 API管理功能被短暂的停止使用.

事故原因: EC2

事故恢复: 4个小时

事故解释:

相关URL: <https://forums.aws.amazon.com/thread.jspa?threadID=17211&start=0&tstart=0>


Amazon的AWS团队的解释是AWS的一些管理软件错误的设置导致了一些客户的虚拟机被误杀.当时为了确保整个AWS服务的安全,AWS团队迅速暂时停止了EC2的管理API功能.

Re: EC2 API outage

 Reply

Posted by:  Peter@AWS

Posted on: Sep 29, 2007 12:04 PM

 in response to: [mediawonder](#)

This is an update on the EC2 issues experienced today. A software deployment caused our management software to erroneously terminate a small number of user's instances. When our monitoring detected this issue, the EC2 management software and APIs were disabled to prevent further terminations. Once we corrected the problem, we restored the management software.

We will contact users that lost instances directly by email. At this point, the service is fully functional, and you should be able to launch replacement instances immediately.

While we have corrected the immediate bug, we are also adding additional checks to prevent this sort of issue from recurring in the future.

We are aware of the following outstanding issues which we are working to resolve now:

1/ Some instances may get stuck in the "shutting down" state until we have completed our clean-up. These instances will not be billed and will be fully terminated shortly.

2/ Some instances will not show their launch indexes in describe-instances API.

We will keep you posted as we resolve these remaining issues.

To address a few of the questions posed on this thread:

The availability of the EC2 APIs is very important and it remains our goal to keep them highly available. We believe disabling the management software was the correct decision because of the risk to running instances. This is not a decision we take lightly, and we will work to avoid having to make this choice in the future.

There was no correlation between the instance terminations, so users with redundancy built into their instance deployments would have been better able to deal with the terminations. We also understand that failure isolation is very important and we are hard at work on additional functionality to help with this.

Please let us know if you experience any unexpected behavior.

The Amazon EC2 Team

3. Feb 15, 2008

08年2月15日, 是Amazon官方对外承认和解释的第一次重大事故. 也从根本的角度影响了产业界对公有云可靠性的认识和警惕. 并直接导致了Amazon决定加强服务可用性的监管和透明化.

事故原因: S3

事故恢复: 3个小时

事故解释:

S3服务子系统的认证(Authentication)服务无法承受突然的大面积的服务请求, 从而导致S3系统瘫痪. AWS的官方解释可参阅:

<http://www.zdnet.com/blog/btl/amazon-explains-its-s3-outage/8010>



For one of our services, the Amazon Simple Storage Service, one of our three geographic locations was unreachable for approximately two hours and was back to operating at over 99% of normal performance before 7 a.m. pst. We've been operating this service for two years and we're proud of our uptime track record. Any amount of downtime is unacceptable and we won't be satisfied until it's perfect. We've been communicating with our customers all morning via our support forums and will be providing additional information as soon as we have it.

Early this morning, at 3:30am PST, we started seeing elevated levels of authenticated requests from multiple users in one of our locations. While we carefully monitor our overall request volumes and these remained within normal ranges, we had not been monitoring the proportion of authenticated requests. Importantly, these cryptographic requests consume more resources per call than other request types.

Shortly before 4:00am PST, we began to see several other users significantly increase their volume of authenticated calls. The last of these pushed the authentication service over its maximum capacity before we could complete putting new capacity in place. In addition to processing authenticated requests, the authentication service also performs account validation on every request Amazon S3 handles. This caused Amazon S3 to be unable to process any requests in that location, beginning at 4:31am PST. By 6:48am PST, we had moved enough capacity online to resolve the issue.

As we said earlier today, though we're proud of our uptime track record over the past two years with this service, any amount of downtime is unacceptable. As part of the post mortem for this event, we have identified a set of short-term actions as well as longer term improvements. We are taking immediate action on the following: (a) improving our monitoring of the proportion of authenticated requests; (b) further increasing our authentication service capacity; and (c) adding additional defensive measures around the authenticated calls. Additionally, we've begun work on a service health dashboard, and expect to release that shortly.

Sincerely, The Amazon Web Services Team

在这次重大宕机之后, AWS团队对业界承诺要作出"Service Health Dashboard", 从而可以透明的使得用户了解AWS各种服务状况.

4. June 5, 2008

08年6月5日, Amazon在东部弗吉尼亚的数据中心找到雷电击.导致该区域的一些EC2服务宕机.

事故原因: 雷电

事故恢复: N / A

事故解释:

雷电导致东部弗吉尼亚的数据中心失去电力.导致EC2宕机.

相关URL:

<http://www.datacenterknowledge.com/archives/2008/06/05/brief-outage-for-amazon-web-service-s/>

Brief Outage for Amazon Web Services

By: Rich Miller
June 5th, 2008



Some web services on Amazon's utility computing platform were briefly disrupted Wednesday afternoon as severe weather caused a power outage near one of the company's data center facilities. At 3:30 pm Eastern time, Amazon said it was "experiencing severe weather near one of our Amazon Web Services locations." A half-hour later the company indicated that the data center had lost grid power, saying that "during the transition to backup generator power, we had a small number of EC2 instances in a single Availability Zone shut themselves down." The instances were restarted quickly, the company said.

Amazon (AMZN) did not identify the location of the data center, but the outage occurred at roughly the same time that severe thunderstorms caused extensive damage in northern Virginia. Amazon has a large data center in Ashburn, Virginia.

5. June 6, 2008

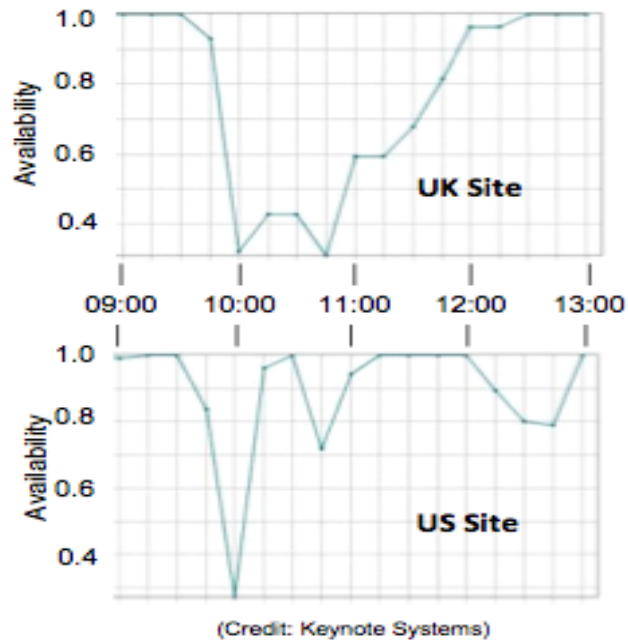
08年6月6日, 基于Amazon本身的网上零售业务突然宕机. 主要是美国和英国的业务. 但AWS本身没有出现异常.

事故原因: Amazon没有对这次事故作出任何官方解释. 只是在非正式场合解释了"Amazon的网络系统非常复杂. 出了点小事情是很偶然和正常的..."

事故恢复: 3个小时

事故解释:

由于Amazon没有对事故作出正式的解释, 业界的猜测是Amazon的负载均衡业务, 例如DNS服务出现了问题. 另外一种说法是Amazon遭到了木马的恶意DDoS攻击. 证据是, 在Amazon主站点宕机的同时, Amazon拥有的IMDB站点(<http://www.imdb.com>)被DDoS通过流量和层7放大攻击. 其攻击流量大概是3Mbits/sec. 下图是当天Amazon美国和英国站点的宕机情况.



6. July 20, 2008

08年的7月20日, S3再次发生重大宕机事故. 许多重要的客户受到影响,例如Twitter. Twitter所有的图像基本上都是存放在Amazon的S3系统里.

事故原因: S3

事故恢复: 8个小时

事故解释: S3服务器之间控制信息流不收敛,从而导致S3服务器无法处理任何用户的服务请求.

同时Amazon也承认EC2的服务也受到了影响. 有些客户的虚拟机无法运行. 另外, Simple Queue Service (SQS)的服务也受到了冲击和中断.

AWS的官方解释为: <http://status.aws.amazon.com/s3-20080720.html>

Amazon S3 Availability Event: July 20, 2008

We wanted to provide some additional detail about the problem we experienced on Sunday, July 20th.

At 8:40am PDT, error rates in all Amazon S3 datacenters began to quickly climb and our alarms went off. By 8:50am PDT, error rates were significantly elevated and very few requests were completing successfully. By 8:55am PDT, we had multiple engineers engaged and investigating the issue. Our alarms pointed at problems processing customer requests in multiple places within the system and across multiple data centers. While we began investigating several possible causes, we tried to restore system health by taking several actions to reduce system load. We reduced system load in several stages, but it had no impact on restoring system health.

At 9:41am PDT, we determined that servers within Amazon S3 were having problems communicating with each other. As background information, Amazon S3 uses a gossip protocol to quickly spread server state information throughout the system. This allows Amazon S3 to quickly route around failed or unreachable servers, among other things. When one server connects to another as part of processing a customer's request, it starts by gossiping about the system state. Only after gossip is completed will the server send along the information related to the customer request. On Sunday, we saw a large number of servers that were spending almost all of their time gossiping and a disproportionate amount of servers that had failed while gossiping. With a large number of servers gossiping and failing while gossiping, Amazon S3 wasn't able to successfully process many customer requests.

At 10:32am PDT, after exploring several options, we determined that we needed to shut down all communication between Amazon S3 servers, shut down all components used for request processing, clear the system's state, and then reactivate the request processing components. By 11:05am PDT, all server-to-server communication was stopped, request processing components shut down, and the system's state cleared. By 2:20pm PDT, we'd restored internal communication between all Amazon S3 servers and began reactivating request processing components concurrently in both the US and EU.

At 2:57pm PDT, Amazon S3's EU location began successfully completing customer requests. The EU location came back online before the US because there are fewer servers in the EU. By 3:10pm PDT, request rates and error rates in the EU had returned to normal. At 4:02pm PDT, Amazon S3's US location began successfully completing customer requests, and request rates and error rates had returned to normal by 4:58pm PDT.

7. June 10, 2009

09年的6月10日, AWS的EC2发生重大宕机事故. 其原因是数据中心遭到雷电击,失去电力.

事故原因: EC2

事故恢复: 8个小时

事故解释:

自然气候, 雷电使得数据中心失去电力.

相关URL:

<http://www.datacenterknowledge.com/archives/2009/06/11/lightning-strike-triggers-amazon-ec2-outage/>

Lightning Strike Triggers Amazon EC2 Outage

By: Rich Miller
June 11th, 2009



9



3



Some customers of Amazon's EC2 cloud computing service were offline for more than four hours Wednesday night after an electrical storm damaged power equipment at one of the company's data centers. The problems began at about 6:30 pm Pacific time, and most affected customers were back online by 11 p.m., according to Amazon's [status dashboard](#). The company said the outage was limited to customers in one of Amazon's four availability zones in the U.S.

"A lightning storm caused damage to a single Power Distribution Unit (PDU) in a single Availability Zone, the company reported. "While most instances were unaffected, a set of racks does not currently have power, so the instances on those racks are down. We have technicians on site, and we are working to replace the affected PDU."

8 July 19, 2009

09年的7月19日, AWS的EC2发生性能和宕机事故.

事故原因: EC2

事故恢复: 2个小时

事故解释: N/A

相关URL:

<http://www.datacenterknowledge.com/archives/2009/07/19/outage-for-amazon-web-services/>

Outage for Amazon Web Services

By: Rich Miller
July 19th, 2009



Amazon's cloud computing services have experienced performance problems this afternoon, with multiple services affected. There are also numerous [reports](#) of users briefly being unable to access the main Amazon.com retail site.

Amazon's [Service Health Dashboard](#) showed problems on the EC2 computing cloud in the US. "We detected a period of elevated packet loss from 12:31 PM PDT to 12:46 PM PDT in a single Availability Zone," Amazon reported. "We are continuing to monitor the situation." The dashboard also showed elevated error rates on its Amazon's CloudFront CDN, SimpleDB database service and Mechanical Turk freelance marketplace. The downtime was also confirmed by monitoring services [CloudStatus](#) and [enStratus](#), which both show the Amazon services available again as of 2 pm Pacific.

The outage is the second in a month for Amazon Web Services, following a June 11 incident in which a [lightning strike](#) damaged power equipment at one of the company's data centers, disrupting service for some AWS customers. Today's problems come at a time of growing scrutiny of the reliability of cloud computing providers. EC2 previously experienced extended outages in [February 2008](#) and [October 2007](#).

9. Oct 5, 2009

09年的10月5日, Bitbucket公司(一个在线开源项目服务公司)在AWS上的业务宕机19个小时.

事故原因: EC2, EBS

事故恢复: 19个小时

事故解释:



Bitbucket在AWS上的服务被黑客用流量攻击的方法打瘫服务. 最开始使用的是UDP Flooding. 然后转换为TCP的Flooding. 服务停顿了19个小时. AWS的运维团队在处理过程中表现的缺乏经验.

相关URL: <http://www.networkworld.com/community/node/45891>

DDoS attack against Bitbucket darkens Amazon cloud

At least 16 hours elapsed before Amazon acknowledged nature of attack, says Bitbucket

By Paul McNamara on Mon, 10/05/09 - 8:53am.

 2 Comments  Print



A crippling DDoS attack over the weekend against open-source hosting service [Bitbucket](#) and Amazon's [EC2 service](#) has questions being raised about the speed and effectiveness of Amazon's response to the emergency, as well as the general reliability of cloud services.

Bottom line for Bitbucket? They're considering switching to a different provider, several of whom were "concerned" enough to offer their services while Bitbucket and Amazon struggled to stem the tide. Bitbucket has almost 19,000 users.

I've requested comment from Amazon's public relations department. (Update: Amazon reply below.)

From a Bitbucket [blog post](#) yesterday by Jesper Noehr:

As many of you are well aware, we've been experiencing some serious downtime the past couple of days. Starting Friday evening, our network storage became virtually unavailable to us, and the site crawled to a halt.

We're hosting everything on Amazon EC2, aka. "the cloud", and we're also using their EBS service for storage of everything from our database, logfiles, and user data (repositories.)

10. Dec 10, 2009

09年的12月10日, AWS的EC2发生宕机事故. 其原因是数据中心遭到雷电击,失去电力. 地点发生在东部北弗吉尼亚的数据中心

事故原因: EC2

事故恢复: 45分钟

事故解释:

自然气候, 雷电使得数据中心失去电力.

相关URL:

<http://www.datacenterknowledge.com/archives/2009/12/10/power-outage-for-amazon-data-center/>

Brief Power Outage for Amazon Data Center

By: Rich Miller
December 10th, 2009



1



Amazon Web Services experienced an outage in one of the East Coast availability zones for its EC2 service early Wednesday due to power problems in a data center in northern Virginia. Failures in a power distribution unit (PDU) resulted in some servers in the data center losing power for about 45 minutes. It took several more hours to get customer instances back online, with all but a “small number” of instances restored within five hours.

“This incident impacted a subset of instances in a single Availability Zone,” said Amazon spokesperson Kay Kinton. “Most of that subset of instances were back online in 45 minutes.”

The issues started at 4 am East Coast time Wednesday, and affected one of the three availability zones in Amazon’s East Coast operation. The zones are designed to provide redundancy for developers by allowing them to deploy apps across several zones.